applied sciences

MDPI

*Article*

# A Deep Reinforcement Learning Design for Virtual Synchronous Generators Accommodating Modular Multilevel Converters

**Mu Yang, Xiaojie Wu * and Maxwell Chiemeka Loveth**

School of Electrical Engineering, China University of Mining and Technology, Xuzhou 221000, China
* Correspondence: xjwu@cumt.edu.cn

**Abstract:** The deep reinforcement learning (DRL) technique has gained attention for its potential in designing "virtual network" controllers. This skill utilizes a novel solution that can avoid the specific parameters and system model required in classical dynamic programming algorithms. However, addressing the issue of system uncertainties and performance deterioration remains a challenge. To overcome this challenge, the authors propose a new control prototype using a twin delayed deep deterministic policy gradient (TD3)-based adaptive controller, which replaces the conventional virtual synchronous generator (VSG) module in the modular multilevel converter (MMC) control. In this approach, an adaptive programming module is developed using a critic fuzzy network point of view to determine the optimal control policy. The modification presented in this framework is able to improve the system stability and resist disruptions while retaining the merits of the conventional VSG control model. The proposed approach is implemented and tested using the DRL toolbox in MATLAB/Simulink.

**Keywords:** virtual synchronous machine (VSG); deep reinforcement learning (DRL); modular multilevel converter (MMC); twin delayed deep deterministic policy gradient (TD3)

check for updates

**Citation:** Yang, M.; Wu, X.; Loveth, M.C. A Deep Reinforcement Learning Design for Virtual Synchronous Generators Accommodating Modular Multilevel Converters. *Appl. Sci.* **2023**, *13*, 5879. https://doi.org/10.3390/app13105879

Academic Editor: Filipe Soares

Received: 4 April 2023
Revised: 28 April 2023
Accepted: 3 May 2023
Published: 10 May 2023

**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

## 1. Introduction

The integration of sustainable energy into the electrical system has been increasing in recent years, leading to a more electrified power grid [1]. Different power electronic integrated equipment could provide the benefits of flexible control and quick reaction. Despite these advantages, weak damping and low inertia can limit broader applications. The stability and security of the power grid are increasingly declining as distributed generation (DG) penetration rates in the grid continue to rise [2].

To address this issue, researchers have proposed using a virtual synchronous machine (VSG) control strategy associated with DRL: a modular multilevel converter (MMC) structure as a control mechanism [3]. The MMC-VSG prototype has gained attention for its potential for application in flexible DC transmission, system stability, power quality, etc. However, it is challenging to balance the steady and dynamic states of traditional MMC-VSG control. Note that it has become urgent to find a way to improve the control structure or to find a novel high-efficiency method to enhance the MMC's performance and robustness.

Efforts have been made to enhance the performance of the VSG control structure. Du et al.'s group explored a virtual inertia frequency control strategy that incorporates a local linearization model of a synchronous motor in the active primary frequency modulation control, in order to better simulate the primary frequency modulation characteristics of the synchronous motor [4]. However, this approach has limitations as the inertia and damping parameters are fixed and cannot be adjusted, as highlighted in previous studies. Another study investigated the output impedance model of the VSG under voltage feedforward

control and evaluated the potential of voltage feedforward control for the VSG based on the current loop [5].

Despite efforts to improve the VSG control structure, existing approaches still face challenges in obtaining precise network characteristics in complex operating conditions of distribution networks [6]. The intelligent algorithm-based "black box" model approach has been investigated as an alternative method for solving such problems. The study in [7] proposed a genetic algorithm photovoltaic virtual synchronous machine (PV-VSG) in comparison with the classic VSG while only considering the active load power change and considering neither the load frequency nor voltage fluctuation, hence, the need for a more detailed study to validate the advantage of the proposed algorithm.

Furthermore, an improved elimination particle swarm optimization (EPSO) technique was encouraged in [8], where the results were based upon data from particle swarm optimization (PSO) and the advocated EPSO. Additionally, no comparison attempt was made against conventional VSG methods, which means that the article failed to acknowledge improvements upon the traditional system. The Back Propagation Neural Network (BPNN) method was recommended in [9], and the fuzzy VSG control rule was introduced in [10].

The knowledge about the reinforcement learning (RL) method associated with voltage stability control is limited among all artificial intelligence (AI) algorithms. RL is a reward-driven mechanism for mapping states to actions that can be used to train an agent using previous data [11,12]. The trained agent controls resources and increases performance by engaging in real time with distribution networks [13]. RL is usually used to solve sequential decision problems and is typically used to formulate a strategy to guide the machine in choosing the best action in different states.

However, it is difficult to obtain training samples containing the state and the optimal action in order to fit the mapping relationship between the state and the action as the optimal strategy. This is due to the complex temporal relationship in the sequence decision problem, that is, the current behavior affects the subsequently received data, and there is a correlation between the data. Under this condition, the method of RL is to let the machine explore the action by itself and then constantly adjust the strategy according to the feedback of the environment.

The ultimate goal is to make the environmental benefits as large as possible [14]. DRL is an RL algorithm that uses a neural network as a parameter structure for optimization [15], which means that DRL is a potent data-driven technique that makes use of the agent's interaction with the dynamic environment to learn the best course of action. On the basis of the action space, DRL settlement can be separated into continuous and discrete control problems [16]. For discrete issues, the traditional deep Q-learning (DQN) approach is appropriate [17], while deep deterministic policy gradient (DDPG) [18] and TD3 are appropriate for continuous problems.

The study in [19] investigated an RL-based adaptive robustness parameter tuning approach for the VSG; however, due to the defect of DQN being used to deal with discrete action problems, the data makes no attempt to engage with the current discussion on continuous action. Although, a control strategy based on DDPG was proposed in [20], allowing for flexible adjustment of the damping and inertia coefficient in response to changing working conditions, the issue of failure brought on by an inflated Q value was not considered.

In the above-mentioned literature related to VSG control strategies, both the improved control structure and intelligent thought methods were given. Few of them considered the TD3 algorithm with regard to the performance of suffering from the active load, frequency, and voltage change, To fill the research gaps of DRL applications in such areas, a TD3-based MMC-VSG structure was developed. In this study, a DRL learning module is employed to replace the VSG module in the control system, which means that the topology of the control system can be much simpler. Moreover, the pre-trained parameter of VSG can be adjusted according to the system change, which will help the system fit into dynamic

circumstances. In view of this, this paper mainly studies the effectiveness and robustness of the DRL learning module and its typical usage in the TD3-VSG control system.

The remaining parts of the study are listed as follows: First, the reasons that cause the contradiction in MMC-VSG's dynamic and steady-state are analyzed. A mathematical model of MMC-VSG and its topology structure is presented, and the VSG control strategy is examined. Secondly, the study focuses on the damping and inertia characteristics affecting the VSG control; the contradiction between J (the moment of inertia) and D (damping coefficient) compels the system to work with the relative "best" working conditions to reconcile transient and steady states. Thus, a DRL-based model, which is proposed in the following section, can separate the dynamic steady state in VSG control.

Thirdly, the Markov decision process (MDP) and the TD3 algorithm are analyzed. The TD3 method is a DRL system developed in light of both the decision-making and perceptual capabilities of deep learning. According to the analysis, a TD3-based MMC-VSG control strategy for replacing the conventional VSG submodule is illustrated. Lastly, based on the prototype, the TD3-based MMC-VSG prototype is built using the DRL toolbox in MATLAB/Simulink software to verify the validity of the raising strategy. Furthermore, a comparison model of DDPG-VSG, fixed-VSG, and vector MMC is also built to demonstrate the effectiveness and superiority.

## 2. Description of MMC-VSG Structure and Modeling

### 2.1. Mathematical Model of MMC

The schematic for the MMC's connection to the AC system is displayed in Figure 1. $R_j$ and $L_j$ indicate the MMC AC side line impedance. $R_0$ is the arm impedance. $i_{vj}$ and $e_{vj}$ denote the output current and voltage. The voltage from the AC grid is $v_{gj}$, and $U_{dc}$ is the DC side voltage, where $j$ stands for phase $abc$.

The main function of MMC is to accomplish the mutual conversion between three-phase AC voltage and DC voltage. One three-phase MMC contains six bridge arms, with two arms existing in each phase unit. Each arm contains N identical submodules in series with half-bridge topology, accompanied by the smoothing inductor $L_0$, which is the inductance of the MMC bridge arm. Figure 1 also depicts the structure of an MMC submodule. Each submodule contains a capacitor ($C_0$) and two complementary IGBTs ($T_1$, $T_2$) linked with reverse continuous flow diodes ($D_1$, $D_2$), respectively. The DC capacitor maintains the DC bus voltage while storing energy.

According to Figure 1, the equations can be derived from the Kirchhoff voltage law (KVL) as

$$\begin{cases} \dfrac{U_{\mathrm{dc}}}{2} = U_{\mathrm{pj}} + R_0 i_{\mathrm{pj}} + L_0 \dfrac{di_{pj}}{dt} + e_{\mathrm{vj}} \\ \dfrac{U_{\mathrm{dc}}}{2} = U_{\mathrm{nj}} + R_0 i_{\mathrm{nj}} + L_0 \dfrac{di_{nj}}{dt} - e_{\mathrm{vj}} \end{cases} \tag{1}$$

The two parts of Equation (1) yield

$$U_{\mathrm{dc}} = U_{\mathrm{pj}} + U_{\mathrm{nj}} + R_0(i_{pj} + i_{nj}) + L_0 \frac{d(i_{pj} + i_{nj})}{dt} = U_{\mathrm{pj}} + U_{\mathrm{nj}} + 2R_0(i_{zj}) + 2L_0 \frac{d(i_{zj})}{dt} \tag{2}$$

$$e_{\mathrm{vj}} = U_{\mathrm{pj}} - U_{\mathrm{nj}} + R_0(i_{\mathrm{pj}} - i_{\mathrm{nj}}) + L_0 \frac{d(i_{\mathrm{pj}} - i_{\mathrm{nj}})}{dt} = U_{\mathrm{pj}} - U_{\mathrm{nj}} + R_0 i_{vj} + L_0 \frac{di_{vj}}{dt} \tag{3}$$

In Equation (2), $i_{zj} = \frac{i_{pj}+i_{nj}}{2}$ stands for the circulating current. Equation (2) can be simplified as

$$U_{\mathrm{dc}} = U_{\mathrm{pj}} + U_{\mathrm{nj}} \tag{4}$$

while the remaining parts are small enough. Under the normal working condition of the MMC, the sum of the lower and upper bridge's input submodule number N is

$$N = N_{pj} + N_{nj} \tag{5}$$

where $N_{pj}$ is the number of submodules of the upper bridge arm in the input state, and $N_{nj}$ is the number of submodules of the lower axle arm in the input state. Equation (5) indicates that half of the submodules in each phase are placed into operation and shall be ensured at any time. The DC side voltage needs the capacitor voltage of N submodules to maintain balance continuously.

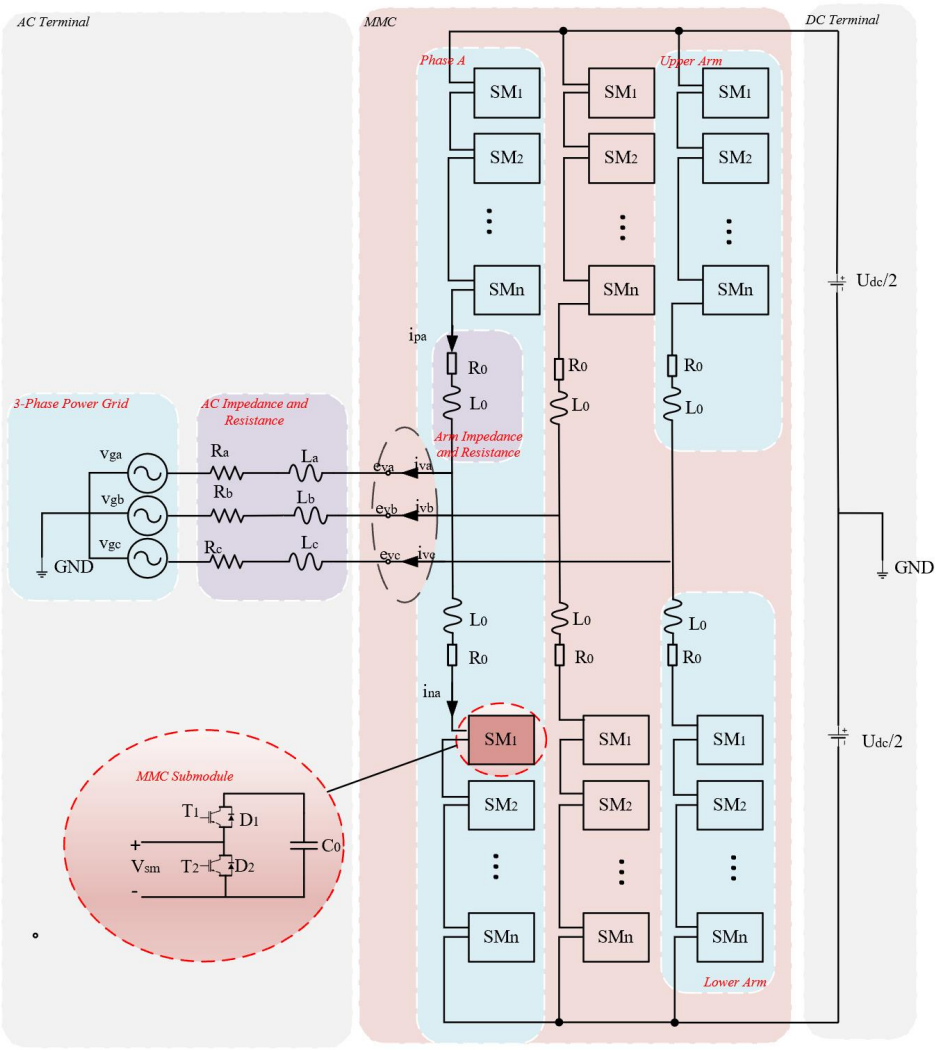



**Figure 1.** The main circuit topology and control method of MMC-VSG.

Table 1 displays the submodule's operating state. The switching function can be summarized as

$$S_{jk} = \begin{cases} 1 & (T_1 on, T_2 off) \\ 0 & (T_2 on, T_1 off) \end{cases} \tag{6}$$

Thus, Equation (5) can be written as

$$N = N_{pj} + N_{nj} = \sum_{k=1}^{N} S_{jkp} + \sum_{k=1}^{N} S_{jkn} \tag{7}$$

according to Equations (4) and (7), the connection between the DC side voltage $U_{dc}$ and the capacitors' voltage of the submodule can be obtained as follows

$$U_{\mathrm{dc}} = \sum_{k=1}^{N_{\mathrm{pj}}} u_{\mathrm{c},p} + \sum_{k=1}^{N_{\mathrm{nj}}} u_{\mathrm{c},n} = \sum_{k=1}^{N} S_{jkp} u_{\mathrm{c},p} + \sum_{k=1}^{N} S_{jkn} u_{\mathrm{c},n} \tag{8}$$

where $u_{c,p}$ and $u_{c,n}$ indicate the submodule's capacitor voltage input by the upper and lower bridge arm, respectively. N submodules in each phase unit are in the input state when the MMC is operating, which means that MMC can output N + 1 level electricity through the corresponding modulation strategy [21]. By changing the on and off distribution relationship between the N submodules on the bridge arm, the anticipated output waveform can be achieved.

**Table 1.** Submodule working status.

| T1 | T2 | Working Statue |
|---|---|---|
| on | off | on |
| off | on | bypass |
| off | off | locked |

### *2.2. MMC-VSG Topology and Control Strategy*

#### 2.2.1. MMC-VSG Structure and Modeling

There are three main components of the SG, which include the prime mover, the synchronous machine, and the power grid [22]. The VSG controls an inverter system that is equivalent to the SG. In contrast with the SG, VSG does not require a complex magnetic field relationship between the stator and rotor; rather, the converter is used to replicate the synchronous motor's exterior features without being constrained by the motor's physical properties. In addition, the electronic parts can enable flexible control of the system. Simultaneously, VSG uses a low-order mathematical model to simplify the high-order motor model, which considerably increases the simulation efficiency by condensing the calculation in the real modeling process. The primary circuit topology and control method of the MMC-VSG is shown in Figure 2.

The MMC control usually adopts a double closed-loop or current inner loop control mode according to the control requirements. The double closed-loop method contains an outer voltage loop and an inner current loop, and the $*$ represents the instruction value for the parameters in the control loop. In previous studies, the inner current loop method was used to replace the voltage loop with virtual stator impedance to calculate the line current. In this study, inner current loop control is manipulated for the VSG, accompanied by nearest level modulation (NLM) to produce the reference modulation signal [23]. As the specification of inner current loop control and NLM modulation have already been extensively discussed in other studies [14,24], this study will mainly focus on providing the VSG-MMC control block diagram instead of going into that in detail.

#### 2.2.2. VSG Control Strategy

As a control method, the VSG enables the inverter to mimic the SG's droop-based operating performance more accurately. The damping and inertia come from the rotor of the conventional SG, so the SG's rotor motion equation is as follows

$$J\omega_0 \frac{\mathrm{d}\omega}{dt} = P_m - P_e - D(\omega - \omega_0) \tag{9}$$

where $\omega$ is output voltage angular frequency for VSG, $\omega_0$ is the rated voltage angular frequency, $P_e$ is the output active power, and $P_m$ is the virtual mechanical power, while $D$ is the virtual damping, and $J$ is the virtual inertia.

The virtual speed controller adopts proportional control.

$$P_m = P_{ref} + k(\omega_0 - \omega) \tag{10}$$

in which $P_{ref}$ is the active power command, and $K$ is the droop coefficient. Therefore, the VSG active power frequency control equation can be obtained by combining Equations (2) and (3)

$$J\omega_0 \frac{d\omega}{dt} = P_{ref} - P_e - D_P(\omega - \omega_0) \tag{11}$$

where $D_P = D + k$ is the equivalent damping coefficient.

Converting Equation (4), we find

$$\frac{\omega - \omega_0}{P_{ref} - P} = \frac{1}{Js\omega_0 + D_P} \tag{12}$$

$$\omega = \frac{P_{ref} - P}{Js\omega_0 + D_P} + \omega_0 \tag{13}$$

It can be seen from Equation (5) that a first-order transfer function makes up the VSG's inertia, and the output power can be obtained from VSG grid-connected circuit, which is

$$P_e = 3\frac{E_g V_g}{X}\sin\delta \approx 3\frac{E_g V_g}{X}\delta = K_P\delta \tag{14}$$

$$Q_e = 3\frac{E_g\cos\delta - V_g}{X}V_g \approx 3\frac{E_g - V_g}{X}V_g \tag{15}$$

where $V_g$ is the grid peak voltage, $E_g$ is the MMC peak output voltage, $X$ is the equivalent impedance of the line, and $K_P = 3\frac{E_g V_g}{x}$. The VSG control block diagram for active power can be achieved from Equations (4) and (7) as shown in Figure 3.

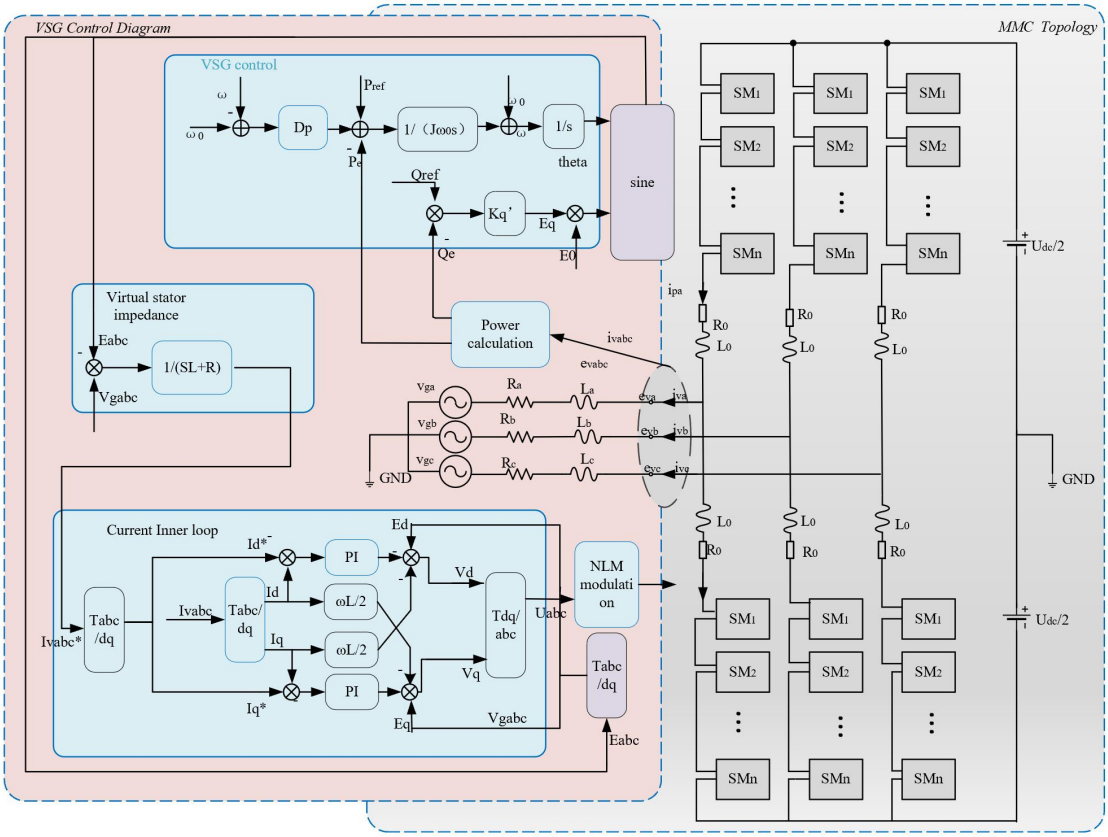



**Figure 2.** The main circuit topology and control framework of MMC-VSG.

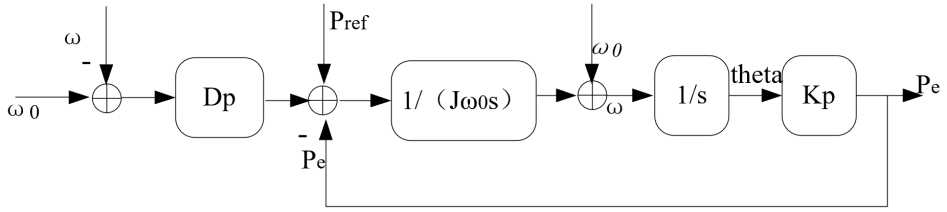



**Figure 3.** Active power control diagram for VSG.

The MMC equivalent AC output port voltage $e_g$ consists of three parts. The first part is MMC-VSG virtual no-load potential $e_0$, which characterizes the terminal voltage of MMC during no-load operation. The second part $e_Q$ corresponds to the part for the reactive power regulation, which can be defined as

$$e_Q = k_q\left(Q_{ref} - Q\right) \tag{16}$$

where $k_q$ is the reactive adjustment coefficient, $Q_{ref}$ is reference reactive power command of the MMC-VSG, and $Q$ is the MMC-VSG instantaneous reactive power.

The third part $e_v$ corresponds to the output of the MMC-VSG voltage-regulating unit and, thus, can be expressed as

$$e_v = k_v\left(E_{ref} - E_g\right) \tag{17}$$

where $k_v$ is the voltage regulation coefficient, and $E_{ref}$ is the MMC reference peak output voltage. We consider that the SG has a primary voltage-regulating curve as shown

$$Q_{\text{ref}} - Q = D_q\left(E - E_{ref}\right) \tag{18}$$

Thus, the MMC-VSG excitation controller mathematical model can be expressed by combining Equations (9) to (11) together

$$e_g = e_0 + e_Q + e_V = e_0 + \left(k_q - \frac{k_v}{D_q}\right)\left(Q_{ref} - Q\right) = e_0 + k_q'\left(Q_{ref} - Q\right) \tag{19}$$

where $k_q' = k_q - \frac{k_v}{D_q}$ can be expressed as the equivalent reactive adjustment coefficient. Thus, according to Equation (12), the reactive power control diagram is shown in Figure 4.

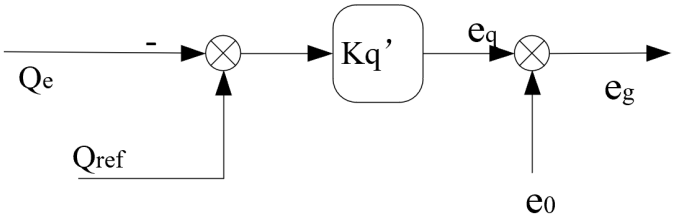



**Figure 4.** Reactive power control diagram for VSG.

From the previous narration, the VSG output active power can be derived from Figure 3,

$$P_e = G_r(s)P_{ref} + G_f(s)(\omega_0 - \omega) \tag{20}$$

where

$$G_r(s) = \frac{k_P}{J\omega_0 s^2 + D_P s + k_P} \tag{21}$$

$$G_f(s) = \frac{k_P(J\omega_0 s + D_P)}{J\omega_0 s^2 + D_P s + k_P} \tag{22}$$

Thus, the output power $P_{\text{out}}$ can be written as

$$\begin{aligned} P_{\text{out}} &= \lim_{s\to 0}(P_e) = \lim_{s\to 0}\Big(G_r(s)P_{ref} + G_f(s)(\omega_0 - \omega)\Big) \\ &= \lim_{s\to 0}\left(\frac{k_P}{J\omega_0 s^2 + D_P s + k_P}P_{ref} + \frac{k_P(J\omega_0 s + D_P)}{J\omega_0 s^2 + D_P s + k_P}(\omega_0 - \omega)\right) \end{aligned} \tag{23}$$

The output power for the steady state is

$$P_{\text{out}} = P_{\text{ref}} + D_P(\omega_0 - \omega) \tag{24}$$

Equation (10) indicates that the steady-state output consists of two parts: active power $P_{ref}$ and steady-state active power deviation $D_P(\omega_0 - \omega)$. Therefore, in order to reduce the steady-state error, $D_p$ should be minimized.

Here, the closed-loop transfer function of the small signal can be summarized as Equation (9), which can be derived as the small signal model of the MMC-VSG shown in Figure 5.

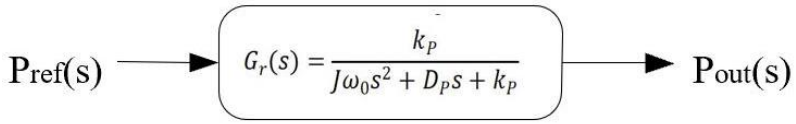



**Figure 5.** Small signal model of the MMC-VSG.

The damping coefficient and natural oscillation angular frequency are as follows

$$\xi = \frac{D_P}{2}\sqrt{\frac{1}{k_P J\omega_0}} \tag{25}$$

$$\omega = \sqrt{\frac{k_P}{J\omega_0}} \tag{26}$$

Equation (14) shows that the system's natural oscillation angular frequency is independent of the value of the damping coefficient $D$ and falls off as the inertia $J$ rises. The system damping is proportional to the damping coefficient $D$ and decreases with the uplift of the inertia $J$.

Figures 6 and 7 show the step response of the MMC-VSG small-signal model under different $D$ and $J$ parameters.

Through the analysis of Figures 6 and 7, it is not difficult to determine that the inertia determines the reaction and system adjustment speed of the dynamic response process, and the $J$ increases with the reaction and system adjustment time. The damping coefficient determines the speed of the oscillation attenuation. The larger $D$ is, the faster the attenuation speed is. Furthermore, the system switches from under-damping to over-damping when $J$ becomes larger or when $D$ becomes smaller.

It can be concluded that the active power steady-state characteristic during the VSG operation is related to $D$, instead of $J$, in accordance with the aforementioned contents and automatic control principle. While the dynamic feature of VSG is connected to $J$ and $D$, an increase in $D$ may lead to growth in the steady-state inaccuracy of the VSG grid connection. The system damping increases with $D$ and reduces the power overshoot, and the system regulation time increases with slower reaction times. Additionally, the system damping decreases as $J$ increases, increasing the overshoot in the dynamic process and degrading stability. Consequently, there are inconsistencies in parameters $J$ and $D$ that prevent them from meeting the demands of a small steady-state error, a large dynamic damping, and a quick system adjustment speed requirement.

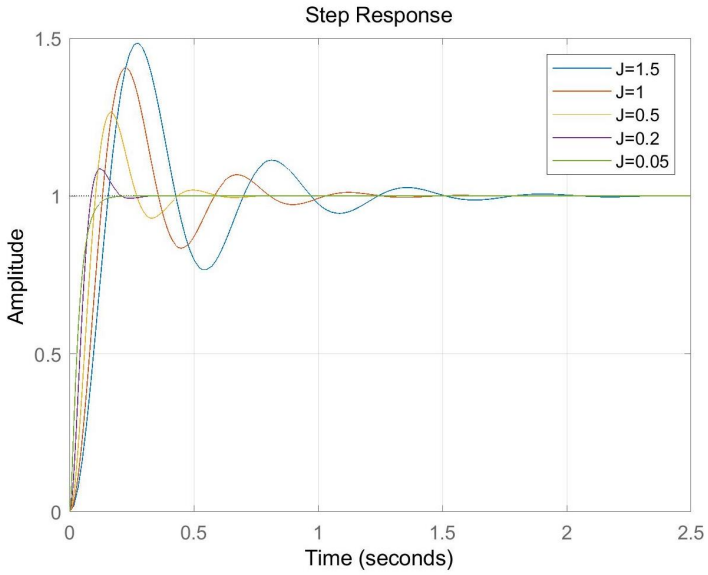



**Figure 6.** Step response for different values of $J$ (with $D = 8$).

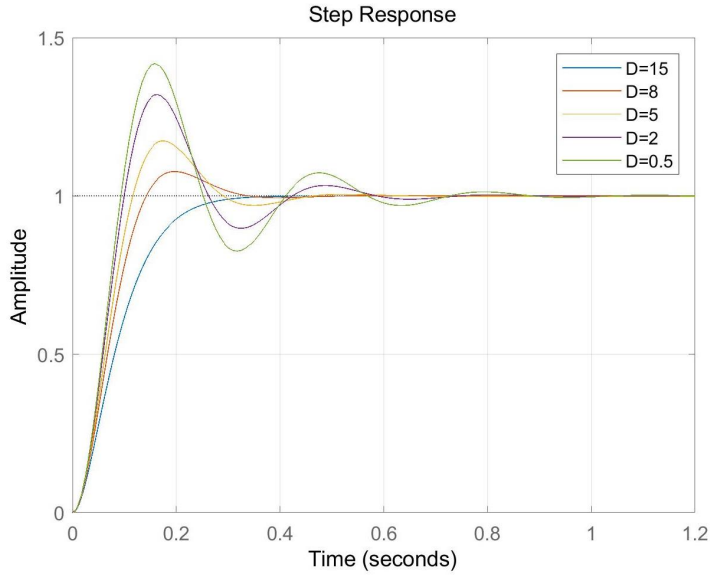



**Figure 7.** Step response for different values of $D$ (with $J = 0.4$).

With the purpose of improving the dynamic features to meet the above properties of the MMC-VSG system without damaging the steady-state characteristics, this article discusses incorporating a VSG control based on the TD3 algorithm.

## 3. Proposed TD3-Based MMC-VSG

### *3.1. Markov Decision Process*

The MDP is used in RL to model the interaction between an agent and the environment, which is depicted in Figure 8. The TD3 technique suggested in this study can also be thought of as an MDP since almost every RL issue can be classified as such. MDP can generally be expressed as <$S$, $A$, $P$, $R$>, where

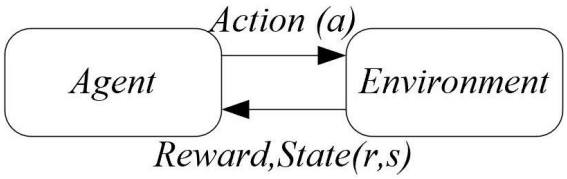



**Figure 8.** The framework of DRL.

$S$ is the name of the state. This includes the environmental conditions at time $t$ examined by the agent. The agent requires proper state information to reasonably manage the VSG control strategy. Here, the state of the active power and reactive power as well as the angular frequency are needed at time $t$. Therefore, the state $S$ is defined as: $S(P_{tn}, Q_{tn}, \omega_{tn})$. $A$ represents the action, it stands for the action at time $t$. In this case, the action can be set as the value of the inertia coefficient ($J$), the damping ($D$), and the voltage ($E$). After each individual action, the next state will be performed by the environment, which can be referred to as $S^{'}$. $P$ is the transfer probability parameter, which is relevant to the uncertainty rate of the system and action value. $R$ is the reward function induced as $R(s, a)$, which is the environment feedback value since the environment implements action $A$ at state $S$.

Thus, under one MDP step, the agent perceives the state $S$ of the environment and takes action $A$ based on the policy $\pi$. Then, the environment transfers to the next state $S^{'}$, and the feedback value $R$ is sent to the agent. According to this feedback, policy $\pi$ will be updated by the agent to maximize the $Q$ function, which is defined as the expected cumulative reward value

$$Q(s, a) = r_{t+1} + \gamma r_{t+2} + \ldots + \gamma^k r_{t+k} = \sum_{k=0}^{\infty} \gamma^k r_{t+k} \tag{27}$$

where $\gamma$ represents the discount rate ($\gamma \in [0, 1]$), which reflects the influence of the future reward value on the current action.

The optimization objective in this paper is to suppress active power oscillation and primary frequency change; hence, the reward function $r$ can be modeled as

$$r_t = -\left( k_1 * {p_{error}}^2 + k_2 * {q_{error}}^2 + k_3 * {f_{error}}^2 + R * \sum_j {u_{t-1}^j}^2 \right) - 100d \tag{28}$$

where $k_1$, $k_2$, $k_3$, and $R$ are coefficient factors, which can be assigned as values of 0.5, 1, 0.5, and 0.01, respectively. $p_{error}$ is the active power error, $q_{error}$ is the reactive power error, $f_{error}$ is the frequency error of the primary frequency modulation, $u_{t-1}^j$ are the previous time step actions, and $d$ is a flag equal to 1 when the emulation is accidentally terminated.

### 3.2. TD3 Algorithm

The TD3 algorithm combines reinforcement learning's capacity for decision making with deep learning's perception [25]. In order to successfully address various flaws in the DDPG algorithm, such as Q-value overestimation, overfitting, and poor updates, TD3-based DRL is applied to create an adaptive control technique.

The TD3 algorithm also builds under the actor–critic framework as does DDPG. TD3 algorithm adopts the method in double Q-learning, which is different from DDPG, which uses two independent neural networks $Q_{\theta 1}$ and $Q_{\theta 2}$ with the network parameters $\theta_1$ and $\theta_2$ to estimate the critic (value). TD3 algorithmindependently emulates two Q-functions, and constructs the target value of critic learning with the smaller Q-value to slow down the overestimation of the critic.

Additionally, each critic network and actor network are accompanied by a corresponding target network $Q^{'}$. The equivalent target networks $Q_{\theta 1}^{'}$, $Q_{\theta 2}^{'}$, and $\pi_{\phi}^{'}$ correspond to the actor neural network $\pi_{\phi}$ and the two critic neural networks $Q_{\theta 1}$ and $Q_{\theta 2}$. The objective of the actor network is to pick up the actions depending on the input variables, while the

critic network is trained to rate the gradient message and equivalent actions given by the actor network, which helps the actor network find a better score.

To implement the TD3 algorithm, first, the target actor network is used to compute action $a^{'}$ under the state $S^{'}$.

$$a' = \pi_{\phi'}(s') \tag{29}$$

Then, adding noise to the target action $a^{'}$ is based on the target policy smoothing regularization

$$\widetilde{a} = \pi'_{\phi}(s') + \varepsilon = a' + \varepsilon \tag{30}$$

where $\varepsilon \sim clip(N(0, \widetilde{\sigma}), -c, c)$ is a clipped noise. The loss function is minimized by Equation (31) with the renovation of two critic neural networks

$$L\left(\theta^{Q_{\theta i}}\right) = \mathbb{E}\left[\left(y - Q_{\theta i}(s,a)^2\right)\right] \tag{31}$$

where the iterative target value $y$ can be derived from the target neural networks to comply with the Bellman equations

$$y = r(s,a) + \gamma \min_{i=1,2} Q_{\theta'_i}(s', \tilde{a}) \tag{32}$$

Using the deterministic policy gradient (DPG) algorithm [26], action network parameters can be updated as follows

$$\nabla_{\phi} J(\phi) = \mathrm{N}^{-1} \sum \nabla_a Q_{\theta_1}(s,a)\Big|_{a=\pi_{\phi}(s)} \nabla_{\phi} \pi_{\phi}(s) \tag{33}$$

The target network and delayed policy update methods are adopted to ensure the training process' validity and stability. The policy network update is delayed due to the actor–critic method's slow parameter update. First, postponing updates might cut down on needless updates that are repeated. However, it is also possible to lessen the faults that gather after numerous upgrades. A soft updating mechanism should also be employed while decreasing the update frequency, which can be expressed as

$$\begin{aligned} \theta'_i &\leftarrow \tau\theta_i + (1-\tau)\theta'_i \\ \phi' &\leftarrow \tau\phi + (1-\tau)\phi' \end{aligned} \tag{34}$$

where $\theta_i^{'}$ is the parameter associated with the target critic network, $\phi^{'}$ is for the target actor network, and ($\gamma \in [0, 1]$ ) is the rate of update.

### 3.3. TD3-Based MMC-VSG Adaptive Control Strategy

Figure 9 illustrates the main steps of the TD3-based MMC-VSG control framework. The training environment includes the entire MMC and power system. Furthermore, the control system is still composed of five parts, which are power calculation, TD3-VSG, virtual stator impedance, current inner loop, and NLM modulation. The red frame shows the detail of TD3-VSG, which replaces the fixed parameter VSG, and the other four parts remain the same.

The AI algorithm can neglect the complex calculation of active and reactive power, and the DRL can change the parameters according to the environment simultaneously, which gives the system better performance during the transient time. The agent observes state $S(P_{tn}, Q_{tn}, \omega_{tn})$ from the environment at time $t$, and then the action $A(J, D, E)$ is produced by the agent according to state $S$. The feedback value $R$ is sent to the agent when the environment moves to the later state $S^{'}$.

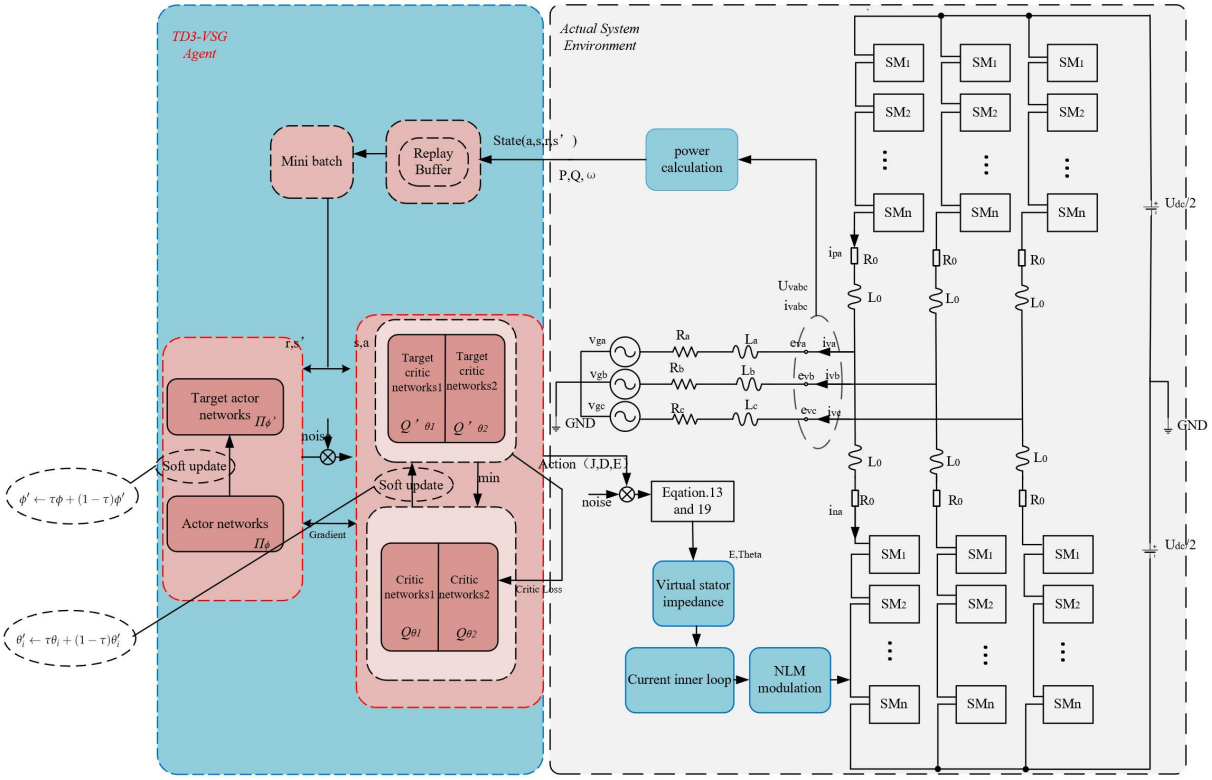



**Figure 9.** TD3-based control framework diagram for MMC-VSG.

## 4. Simulation Results

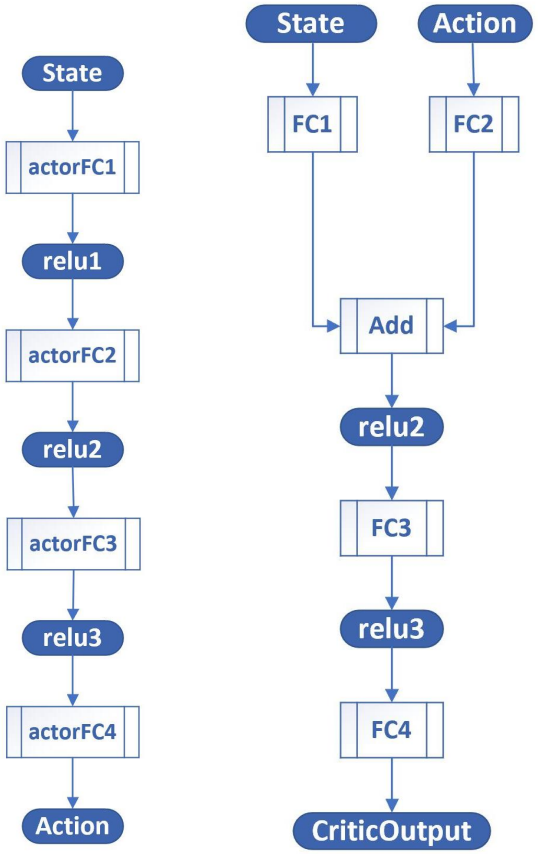
### *4.1. Training of TD3*

A VSG-MMC model on the MATLAB/Simulink platform is implemented to serve as a testing system for the purpose of determining whether the suggested DRL-based agent is effective. Figure 10 shows the layer graph of the critic/actor network proposed by the article.

Both the actor and critic networks have the same size with five layers, which are one input layer, one output layer, and three hidden layers. Each hidden layer is formed from a non-saturated function. It is possible to think of TD3 training as an iterative process of finding the best possible control strategy. The goal of the TD3 training in this research is to create an agent that can provide the best MMC-VSG parameter settings under various damping and inertia conditions on the MATLAB/Simulink platform, the parameters of which are displayed in Table 2.

**Table 2.** Parameters of the TD3-VSG training setting.

| Simulation Parameter | Value |
|---|---|
| Training Episode | 600 |
| Discount Factor | 0.995 |
| MiniBatch Size | 512 |
| Learning Rate (Actor) | 0.0001 |
| Learning Rate (Critic) | 0.001 |
| Soft Update Discount Factor | 0.01 |




(**a**) (**b**)

**Figure 10.** Layer graph. (**a**) Actor network. (**b**) Critic network.

Figure 11 depicts the average rewards with regard to training iterations using the framework developed in Section 3. In the same picture, a DDPG comparison technique with identical parameters is also displayed. At the training stage, the convergence patterns of the two methods are very similar. The blue/red star represents every 10 episodes of average rewards while training, whereas the line represents the reward $R$'s tendency from beginning to finish. During the first 100 training sessions, the TD3 reward amount varies between $-400$ and $-600$, while the DDPG reward value varies between $-400$ and $-500$, and the matching output changes greatly.

The TD3-VSG's reward amount grows considerably after 200 episodes, in contrast with the DDPG-VSG, which retains the value of a modest reward until roughly 240 training sessions. The TD3 algorithm employs the approach of double Q-learning, resulting in a quicker convergence speed. As the number of training episodes increases, the reward value eventually converges between $-30$ and 0, indicating that the agent has adopted the best control strategy.

Table 3 compares the training time for both methods. The training time for the TD3 algorithm was 14.7 h compared to 11.2 h for the DDPG method.

**Table 3.** The comparison experiment parameters.

| Parameter | TD3 | DDPG |
|---|---|---|
| training Time | 14.7h | 11.2h |

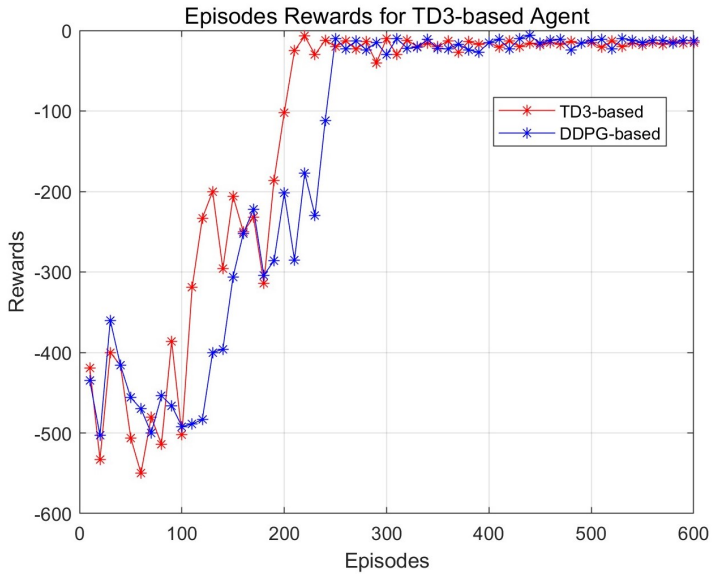



**Figure 11.** Episode reward for TD3 agent training.

### *4.2. Reliability and Effectiveness Experiments*

An MMC grid-connected system bottom on the VSG emulation model was created in MATLAB/Simulink to validate the dependability and efficacy of the proposed prototype for optimizing specific parameters by integrating the simulation results to examine the dynamics and reaction times of the system.

The simulation time of each episode was set to 10 s before the training, and the sampling time was set to 0.0005 s. Additionally, the random disturbance must be added to the environment to aid the model in determining the best action decision-making mechanism and assessing the model's robustness. The parameters for the upcoming experiment are listed in Table 4.

**Table 4.** The comparison experiment parameters.

| Simulation Parameters | Symbol | Value |
|---|---|---|
| Simulation Time | $T_s$ | 10 s |
| Sampling Time | $T_f$ | 0.0005 |
| Fixed VSG Inertial | $J_{MMC}$ | 0.4 |
| Fixed VSG Damping | $D_{MMC}$ | 8.4 |
| DC Side Voltage | $U_{dc}$ | 800 V |
| Grid Line-Line Voltage | $E_{ab}$ | 380 V |
| System Rated Frequency | $f$ | 50 Hz |
| Sub-Modules of MMC Bridge | / | 17 |
| Sub-Module Capacitance | $C_{SM}$ | 60 uf |
| PI Parameter for Current Inner Loop | $K_p/K_i$ | 180/3200 |
| Arm Inductance | $L_0$ | 1.5 mH |
| Arm Resister | $R_0$ | 0.03 Ω |
| AC Side Inductance | $L_i$ | 2.5 mH |
| AC Side Resister | $R_i$ | 0.05 Ω |

The MMC has different structural properties from other inverters. It comprises several submodules in each arm. The dispersion and volatility of the submodule capacitor voltage are the main determinants of whether or not the MMC can run steadily. The output voltage is displayed in Figure 12 when the MMC is off grid. The MMC has 17 output levels, which complies with the 2N + 1 rule.

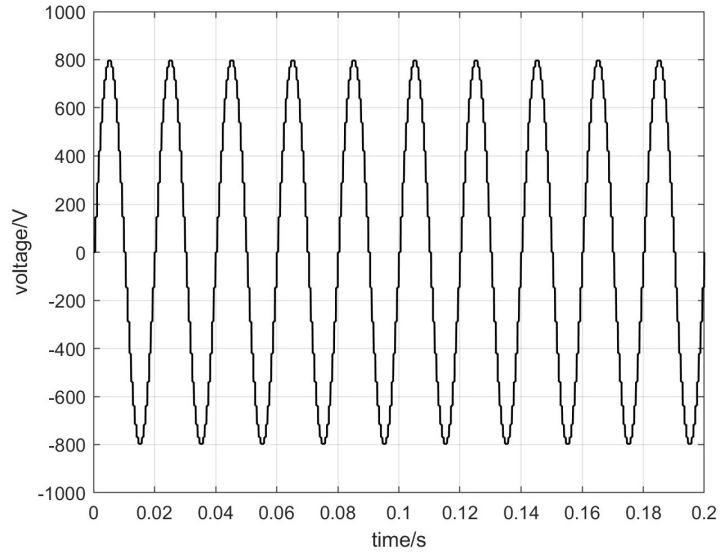



**Figure 12.** MMC output voltage (off grid).

Each phase of the MMC has a lower and an upper bridge arm with eight identical submodules. The bridge arm of each phase is shown by their actual measured capacitor voltage values in Figure 13. It can be concluded that the typical capacitor voltage sequencing mechanism in NLM is used for the submodule capacitor voltage balancing. As a priority, the MMC will charge the capacitor submodule with the lowest voltage level and discharge that with the highest voltage level. The capacitor voltage swings from 95 to 110 V can be seen in the waveform, which satisfies the criteria for steady operation of the MMC. Simulations in this part also demonstrate the success of the modulation method and the functionality of TD3-based VSG-MMC.

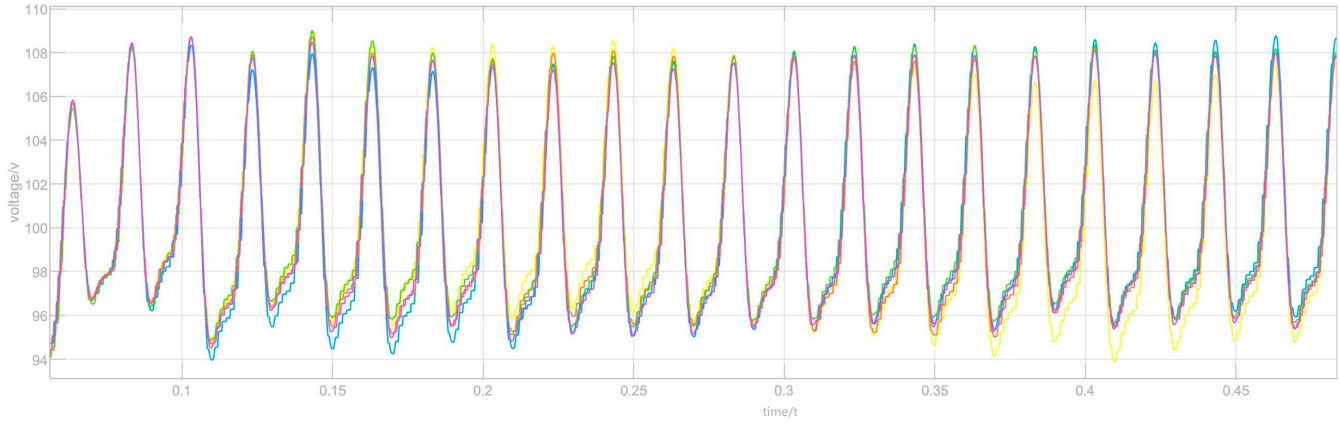



**Figure 13.** The submodule capacitor voltage in one arm.

### *4.3. Performance Evaluation*

The results of the experiment were analyzed to demonstrate the significance of the suggested TD3-VSG approach. Two parts were set to showcase the effectiveness of the method. The first part was to compare four MMC control methods under a change in the load's active power, while the second part aimed to validate the TD3-VSG approach's proficiency in robust performance. Four controllers were compared in this scenario, the DDPG-based controller, the fixed VSG controller in conventional control algorithms, the vector MMC controller, and the TD3-based controller.

4.3.1. A Comparison of the Experiments Performed Under Active Power Change

The results of four controllers in active power and frequency change are drawn as follows.

The results of the four controllers' active power and frequency changes are illustrated in Figures 14 and 15. Figure 14 shows that the TD3 model had the fastest response speed and the smallest turbulence compared to the other two MMC-VSG methods. The fixed VSG model reached a peak of around 360 kW, while the DDPG method reached 310 kW, and the TD3 method reached 290 kW, with the transient time decreasing in the same sequence. All three VSG control methods had a steady active power of around 280 kW, in contrast with the vector MMC method, which showed a relatively shorter transient state with output power remaining after oscillation.

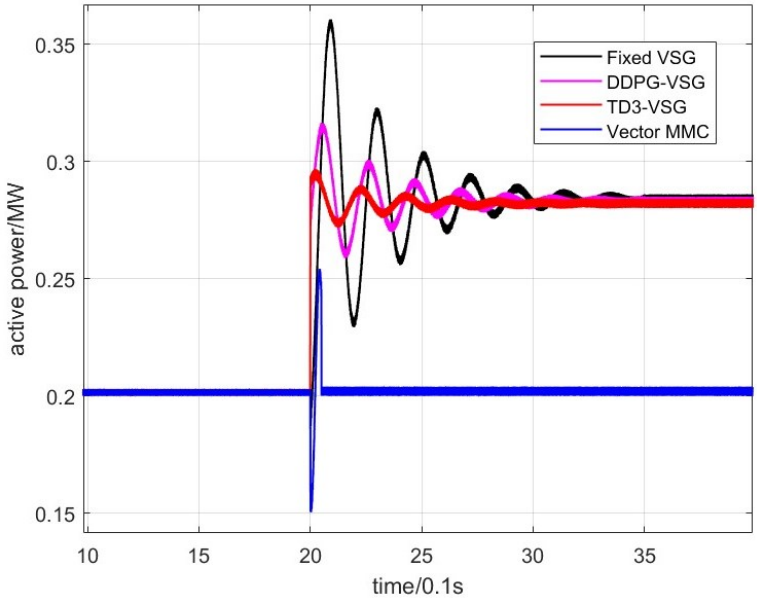



**Figure 14.** A comparison of the results of a change in active power.

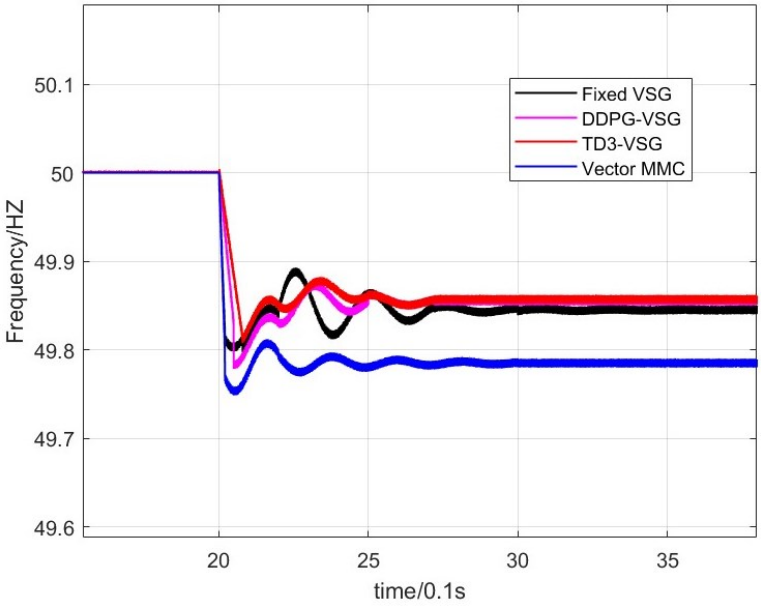



**Figure 15.** A comparison of the results of a change in frequency.

In the experiment, voltage changes were observed in response to a sudden increase in the grid-side load. The excitation controller in the MMC-VSG system adjusted its power output to maintain the grid voltage and frequency stability, resulting in optimal stability and response times. However, when the grid-side load increased rapidly, the system's output power remained unchanged, and the vector control method exhibited only a short transient process. In this scenario, the MMC system was unable to independently adjust its power output based on the change in grid-side load, leading to a lack of fast power response due to its lack of inertia and damping characteristics as well as its inability to participate in power-grid regulation.

The performance of the four control methods was further analyzed by comparing their frequency change under grid-connected operating conditions as shown in Figure 15. All three VSG control methods showed similar patterns, ending at around 49.85 Hz, while the vector MMC ended at 49.79 Hz. The vector MMC showed a larger frequency deviation due to the PLL's poor ability to track the grid frequency after a drop. During the dynamic response period, the TD3 model showed the smallest frequency fluctuation and dynamic response time among the three VSG methods.

Hence, it can be concluded that the TD3 technique has stronger resistance to disturbance, better dynamic performance, and a quicker response time compared to the DDPG-based algorithm and fixed VSG control. The stability of the training environment was improved, and the overestimation error was reduced through the use of a truncated end naturalized Q-network critic, which minimized the overestimation bias of a single agent's critic network in the population. The algorithm's performance was also improved, indicating that the recommended method has superior demonstration and robustness against uncertainty in conditions.

The simulation results for the output current and voltage are shown in Figure 16. It can be observed that a sudden increase of 200 kW in the grid active power causes the MMC system's output voltage to decrease significantly, with a decrease in amplitude to around 230 V. This is due to the use of a fixed number control with the vector control, causing the current to increase significantly from 70 to around 90 A in order to maintain constant active power. Both the current and voltage have a very short transient time. The other three sub-figures demonstrate that the VSG control can become involved in the regulation of both the voltage and frequency of the AC grid.

The voltage and current vary in response to changes in power, with TD3-VSG having the smallest distortion of about 140 A in current compared to the largest distortion of about 100 A for the fixed VSG. The dynamic response time for all three methods is less than 1 s. Compared to the vector control, the dynamic response time during the fluctuation process is significantly longer for the VSG system, and the steady-state amplitude will also change. As a result, under the influence of a rapid load increase, the system's current varies and becomes larger than it was prior to the load increase after stability. The voltage amplitude remains largely unchanged during this time.

However, the VSG system successfully restrains the system's output voltage from decreasing greatly, thereby stabilizing the voltage. The TD3-based system outperforms the other two VSG algorithms and brings the properties of damping and inertia into the control of the MMC, allowing the output voltage and frequency to be self-adaptive in response to changes in load. This significantly increases the power system's capacity to operate stably. In conclusion, when compared to the vector control, the TD3 control does not result in a significant reduction in voltage amplitude and demonstrates better performance and robustness against system uncertainty.

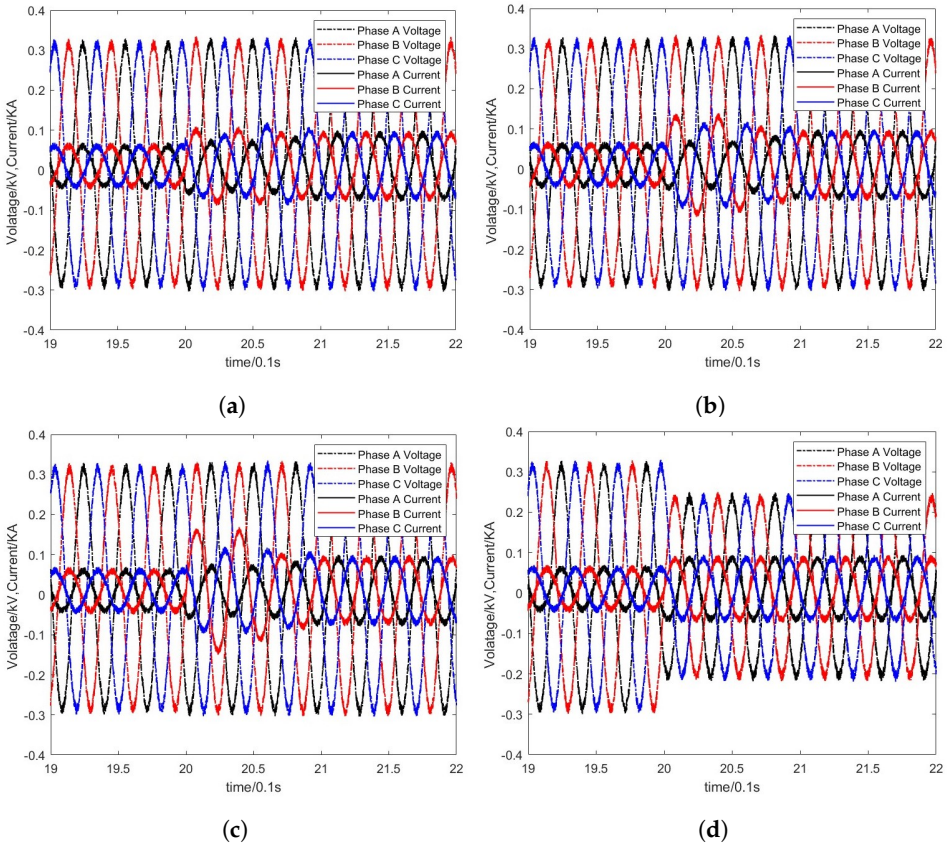



**Figure 16.** Voltage and current change. (**a**) TD3-VSG. (**b**) DDPG-VSG. (**c**) Fixed VSG. (**d**) Vector MMC.

#### 4.3.2. Robustness Evaluation

To further establish the robustness of the proposed model, a comparative experiment was conducted to evaluate the impact of parameter mismatch on the total harmonic distortion (THD) in the output current. The robustness of the proposed TD3-VSG was assessed by varying the input inductance value and input resistance value of the MMC arguments as illustrated in Figure 17. The assessment involved applying a 2 to 3 mH ($0.8L_i$ to $1.2L_i$) parameter mismatch for AC load inductance and 0.04 to 0.06 Ω ($0.8R_i$ to $1.2R_i$) mismatch for resistance, while utilizing the MMC parameters listed in Table 3 throughout the robustness evaluation.

The assessment interval was 10 percent for both parts. Figure 17c shows that the performance of the fixed MMC-VSG approach degraded significantly as the load inductance decreased from 2.5 to 2 mH ($L_i$ to $0.8L_i$) due to its inaccuracy accompanied by parameter changes. Conversely, the proposed TD3-based solution was capable of mitigating performance degradation caused by parametric uncertainties without requiring a physical or mathematical model of the system, in contrast to vector control MMC and fixed MMC-VSG control schemes. Furthermore, according to Figure 17, the TD3 control technique's performance was minimally affected by both incremental and decremental changes in the MMC parameters.

This indicates that the proposed module is capable of maintaining acceptable output current, voltage, and power levels while following their set-up values under parameter variations with restricted steady and transient errors. Thus, based on the above discussion, it can be concluded that our test proves that the raising prototype is robust under both dynamic and steady operating situations, thereby demonstrating the applicability of our proposal and its ability to tolerate an extensive range of uncertainties while retaining excellent performance.

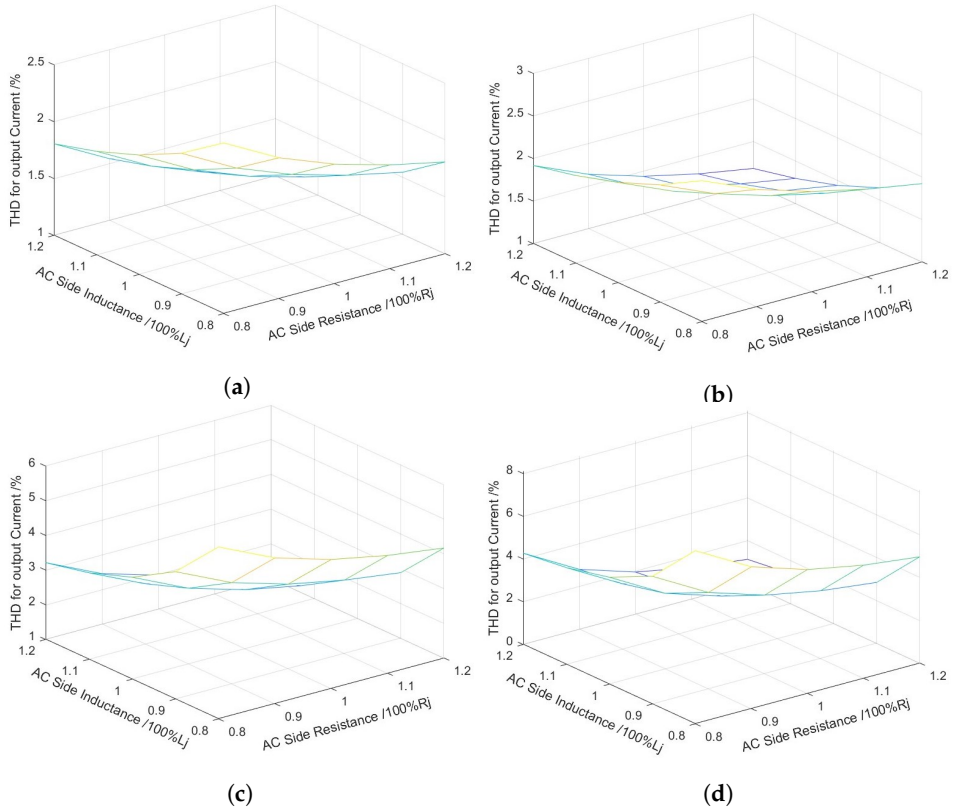



**Figure 17.** THD for output current $i_{vj}$ with mismatch L and R. (**a**) TD3-VSG. (**b**) DDPG-VSG. (**c**) Fixed VSG. (**d**) Vector MMC.

## 5. Conclusions

To mitigate the limitations of the vector MMC control and traditional VSG-MMC control, this study investigated the feasibility of a novel TD3-based adaptive control strategy for optimizing the MMC-VSG system. To be more specific, this approach can effectively address the challenges posed by external disturbances and parametric uncertainties. Our proposed TD3-VSG prototype facilitates stable and rapid learning and reimburses system model indeterminacy, thereby improving the system robustness and performance against disturbance.

In addition to suppressing active power oscillation and primary frequency modulation, we developed a preliminary definition of the TD3-VSG algorithm and established its general framework. After training, the algorithm enhanced the system's transient performance and operational stability, while also improving the voltage amplitude performance and system constancy relative to other control strategies.

Experimental results based on the MMC demonstrated the benefits of our proposed solution, which shows promise for fulfilling control tasks. However, due to inevitable discrepancies between emulation models and real-time dynamics, the consistency and attenuation of TD3-VSG in power electronics remain as research directions that require further attention for industrial implementation.

Potential future studies on this subject will include extending the reward functions for various parameters, such as line loss, harmonic distortion rate, and submodule capacitance balance. Additionally, a multi-terminal MMC project can also be established to emulate a high-voltage direct current (HVDC) system.

**Author Contributions:** Writing—original draft, M.Y.; Writing—review & editing, M.C.L.; Supervision, X.W. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the National Key Research and Development Program of China under Grant 2017YFF0210600.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Lv, Z.; Zhong, Q.C. Control of modular multilevel converters as virtual synchronous machines. In Proceedings of the IEEE Power & Energy Society General Meeting (PESGM), Chicago, IL, USA, 16–20 July 2017; pp. 1–5.
2. Wei, W.; Lin, J.; Yi, C.; Yong, L. A parameter alternating VSG controller of VSC-MTDC systems for low frequency oscillation damping. *IEEE Trans. Power Syst.* **2020**, 35, 4609–4621. [CrossRef]
3. Zhong, Q.; Weiss, G. Synchronverters: Inverters that mimic synchronous generators. *IEEE Trans. Ind. Electron.* **2011**, 58, 1259–1267. [CrossRef]
4. Du, W.; Jiang, Q.; Chen, J. Virtual inertial frequency control strategy for micro grid power supply. *Power Syst. Autom.* **2011**, *35*, 26–31.
5. Yang, Y.; Chen, J.; Wan, Y.; Zhang, X.; Wang, K. Impedance modeling and grid connection stability analysis of grid voltage feedforward control VSG. *Electr. Power Eng. Technol.* **2021**, *40*, 52–61.
6. Bagheri, P.; Xu, W. Model-free volt-var control based on measurement data analytics. *IEEE Trans. Power Syst.* **2019**, *34*, 1471–1482. [CrossRef]
7. Rehman, H.U. Study on VSG Stability Control and Primary Frequency Modulation of Renewable Energy Generation Unit. Ph.D. Dissertation, North China Electric Power University, Beijing, China, 2021. [CrossRef]
8. Cheng, J.; Xiao, J.; Wang, X. VSG control strategy based on improved particle swarm optimization algorithm. *Control Eng.* **2021**, *28*, 20282037. [CrossRef]
9. Wanigasekara, C.; Swain, A.; Nguang, S.K.; Prusty, B.G. Improved Learning from Small Data Sets Through Effective Combination of Machine Learning Tools with VSG Techniques. In Proceedings of the 2018 International Joint Conference on Neural Networks (IJCNN), Rio de Janeiro, Brazil, 8–13 July 2018; pp. 1–6. [CrossRef]
10. Zheng, J.; Wang, A.; Zhang, T.; Zhang, H.; Gu, N. Improved Adaptive Control Strategy of Inertia for Virtual Synchronous Generator Based on Fuzzy Control. In Proceedings of the 2019 Chinese Control Conference (CCC), Guangzhou, China, 27–30 July 2019; pp. 2531–2535. [CrossRef]
11. Schultz, W. Neuronal reward and decision signals: from theories to data. *Physiol. Rev.* **2015**, *95*, 853–951. [CrossRef] [PubMed]
12. Cao, D.; Hu, W.; Xu, X.; Wu, Q.; Huang, Q.; Chen, Z.; F Blaabjerg. Deep reinforcement learning based approach for optimal power flow of distribution networks embedded with renewable energy and storage devices. *J. Mod. Power Syst. Clean. Energy* **2021**, *9*, 1101–1110. [CrossRef]
13. Zhang, D.; Han, X.; Deng, C. Review on the research and practice of deep learning and reinforcement learning in smart grids. *CSEE J. Power Energy Syst.* **2018**, *4*, 362–370. [CrossRef]
14. Sutton Richard, S.; Barto Andrew, G. *Reinforcement Learning: An Introduction*, 2nd ed.; MIT Press: Cambridge, MA, USA, 2018.
15. Liu, Q.; Zhai, J.; Zhang, Z.; Zhong, S.; Zhou, Q.; Zhang, P.; Xu, J. A Survey on Deep Reinforcement Learning. *Chin. J. Comput.* **2018**, *41*, 1–27. [CrossRef]
16. Arulkumaran, K.; Deisenroth, M.P.; Brundage, M.; Bharath, A.A. Deep reinforcement learning: A brief survey. *IEEE Signal Process. Mag.* **2017**, *34*, 26–38. [CrossRef]
17. Mnih, V.; Kavukcuoglu, K.; Silver, D.; Rusu, A.A.; Veness, J.; Bellemare, M.G.; Graves, A.; Riedmiller, M.; Fidjeland, A.K.; Ostrovski, G.; et al. Human-level control through deep reinforcement learning. *Nature* **2015**, *518*, 529–533. [CrossRef] [PubMed]
18. Lillicrap, T.P.; Hunt, J.J.; Pritzel, A.; Heess, N.; Erez, T.; Tassa, Y.; Silver, D.; Wierstra, D. Continuous Control with Deep Reinforcement Learning. 2019. Available online: http://arxiv.org/abs/1509.029711 (accessed on 7 November 2021).
19. Wu, W.; Guo, F.; Ni, Q.; Liu, X.; Qiu, L.; Fang, Y. Deep Q-Network based Adaptive Robustness Parameters for Virtual Synchronous Generator. In Proceedings of the 2022 IEEE Transportation Electrification Conference and Expo, Asia-Pacific (ITEC Asia-Pacific), Haining, China, 28–31 October 2022; pp. 1–4. [CrossRef]
20. Xiong, K.; Hu, W.; Zhang, G.; Zhang, Z.; Chen, Z. Deep reinforcement learning based parameter self-tuning control strategy for VSG. *Energy Rep.* **2022**, *8*, 219–226. [CrossRef]
21. Bo, C.; Fan, Y.; Kaicheng, H. Optimization of MMC Voltage Equalization Control Parameters. *J. Electron. Meas. Instrum.* **2019**, *33*, 19–24.
22. Li, J. Characteristics of Synchronous Generator. In *Design and Application of Modern Synchronous Generator Excitation Systems*; IEEE: Piscataway, NJ, USA, 2019; pp. 35–66. [CrossRef]

23. Zhou, W.; Dong, Y.; Yang, H.; Li, W.; He, X.; Hu, J.; Yuan, X. Common-mode voltage injection-based nearest level modulation with loss reduction for modular multilevel converters. *IET Renew. Power Gener.* **2016**, *10*, 798–806. [CrossRef]
24. Zheng, X. *Flexible High Voltage Direct Current System*; China Machine Press: Beijing, China, 2013.
25. Fujimoto, S.; Hoof, H.; Meger, D. Addressing Function Approximation Error in Actor-Critic Methods. In Proceedings of the International Conference on Machine Learning, Stockholm, Sweden, 10–15 July 2018.
26. Silver, D.; Lever, G.; Heess, N.; Degris, T.; Wierstra, D.; Riedmiller, M. Deterministic policy gradient algorithms. In Proceedings of the International Conference on Machine Learning, Beijing, China, 21–26 June 2014; pp. 387–395.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.